

FREETALKY: Don't Be Afraid! Conversations Made Easier by a Humanoid Robot using Persona-based Dialogue

Chanjun Park^{1,2*}, Yoonna Jang^{1*}, Seolhwa Lee^{3*}, Sungjin Park^{4*}, and Heuseok Lim^{1†}

¹Korea University

²Upstage

³University of Copenhagen

⁴NAVER Corp.

{bcj1210, morelychee, limhseok}@korea.ac.kr

chanjun.park@upstage.ai

sele@di.ku.dk

sungjin.park@navercorp.com

Abstract

We propose a deep learning-based foreign language learning platform, named FREETALKY, for people who experience anxiety dealing with foreign languages, by employing a humanoid robot NAO and various deep learning models. A persona-based dialogue system that is embedded in NAO provides an interesting and consistent multi-turn dialogue for users. Also, an grammar error correction system promotes improvement in grammar skills of the users. Thus, our system enables personalized learning based on persona dialogue and facilitates grammar learning of a user using grammar error feedback. Furthermore, we verified whether FREETALKY provides practical help in alleviating xenoglossophobia by replacing the real human in the conversation with a NAO robot, through human evaluation.

Keywords: FreeTalky, Grammar Error Feedback, Humanoid Robot NAO, Persona Dialogue

1. Introduction

Following multiculturalism and hyper-connectivity, many societies value multilingual skills as a demonstration of the ability to work effectively with customers, clients, and businesses of various countries and cultures (Damari et al., 2017). In this context, native-speaking classes and foreign language phone calls for foreign language learning are available in most countries.

However, the costs are very high and some learners experience a mental block in learning foreign languages (*i.e.*, foreign language anxiety, also known as xenoglossophobia). It can be described as a feeling of uneasiness, worry, nervousness, and apprehension that is experienced while learning or using a second or foreign language (Horwitz et al., 1986; Donmus, 2010).

To address these issues, we propose FREETALKY, a software platform that allows users to communicate comfortably as if they were conversing with a native speaker by conversing with a robot instead through realistic free talking. With FREETALKY, users obtain a learning effect beyond the traditional classroom learning methods, by experiencing direct conversations without the mental pressure, rather than by only seeing and hearing.

The FREETALKY platform combines speech recognition, a persona-based dialogue system, English grammar error correction, and a humanoid robot NAO,

for English learning that is user-friendly and practically helpful, by reflecting user perspectives. Specifically, the persona-based dialogue system provides customized multi-turn dialogue, while the English grammar error correction software provides customized grammar correction. These features allow users to have user-friendly conversations with the robot, learn foreign languages effectively, and even gain assistance with their xenoglossophobia. The contributions of this study are as follows.

- To the best of our knowledge, FREETALKY is the first deep learning-based conversation platform for alleviating xenoglossophobia.
- The persona-based conversation system in FREETALKY allows users to customize foreign language learning and provides practical feedback on grammatical errors through a deep neural network-based grammar error correction software.
- Through an evaluation and survey of real users of FreeTalky, the validity and feasibility of FreeTalky were verified, and various experimental empirical analyses and error analyses were performed.

2. Related Work

To the best of our knowledge, there is no foreign language learning software for alleviating xenoglossophobia. In fact, a majority of the commercialized software products for foreign language learning are developed

*These authors contributed equally.

†This author is the corresponding author.

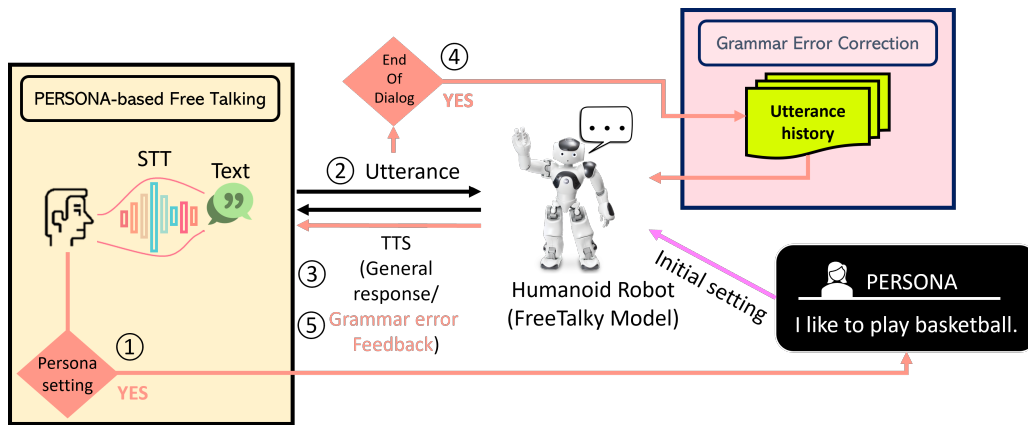


Figure 1: Overview of FREETALKY process.

for the general public, with the primary objective of improving English language skills, and do not consider xenoglossophobia.

There are three major foreign language learning platforms: Andy ¹, Cake ², and SpeakNow ³. However, despite the commercialized service system, there is considerable room for improvement.

First, Andy does not provide contextual dialogue, and users cannot receive evaluation or feedback on their learning. Second, Cake provides feedback on pronunciation and expressiveness; however, there are no means to have a natural conversation through an actual dialogue system. Additionally, it only provides traditional evaluation methods, such as word quizzes and fill in the blanks. Further, there is no feature that provides practical feedback for English grammar. Finally, SpeakNow, an educational English conversation application using 3D avatars, provides evaluation of grammar accuracy, expressiveness, and pronunciation, exploiting various conversation situations and users' voices. However, it is not a type of free conversational model, as it only provides a service that selects and answers a given sentence, or reads the given sentences.

In light of these limitations, we propose FREETALKY, which is equipped with a persona conversation system that enables personalized learning and an English grammar corrector that provides practical learning feedback.

3. FREETALKY

Workflow of FREETALKY In this section, we describe the overall workflow of FREETALKY, as shown in Figure 1. The process is explained step by step as follows.

1. The user sets the persona of the robot (e.g., (1) [*I like to play basketball., ...*], (2) [*I have two daughters., ...*], (3) [*I am a graduate student., ...*] by se-

lecting a number. Each number is assigned a randomly chosen list of personas, and this information is entered into the system. Automatic persona setting is provided for the convenience of the user; however, the users can set the persona of the robot themselves as well.

2. The user begins the conversation with the NAO robot to practice speaking in English. The input of the user is further converted into text using the speech-to-text (STT) module, and passed on to the dialogue system. The dialogue model inputs the converted text with persona sentences and previous history of the dialogue. Next, the model generates a response.
3. The generated response is converted into speech using the text-to-speech (TTS) module of the NAO robot, for it to provide the answers to the user. Through this process, the user is able to receive the output from the model in the form of human voice.
4. When the conversation is completed by the end command (i.e., *Bye.*), all utterances are transmitted to the grammar error correction (GEC) module, where the grammatical errors in the utterances are detected.
5. The robot provides feedback on the sentences that include grammatical errors, if there are any. By receiving feedback, the users learn English grammar, in addition to experiencing a seamless conversation in English.

NAO Robot According to shamsudin2011humanoid, NAO is a humanoid robot that has been purposely designed to seem approachable. We embedded our FREETALKY in the NAO robot to allow the robot to be able to communicate with people more easily and effectively. By utilizing this humanoid robot as an educational assistant, people who are

¹<https://andychatbot.com/>

²<https://mycake.me/>

³<https://www.speaknow.ai>

learning the English language would experience more natural conversations, rather than studying alone.

Speech-to-Text We adopted an off-the-shelf STT module known as CLOVA speech recognition (CSR) (Ha et al., 2020) API of Naver. It obtains the speech input of the user and converts it into text. Further, it sends the converted text to the dialogue system of the model.

Text-to-Speech The TTS system is based on ALText-ToSpeech, which is a built-in TTS system of the NAO robot. This module is used for providing feedback to the users on their grammatical errors. In addition, the robot alters its emotion and voice, which allows the users to experience a life-like conversation with foreigners.

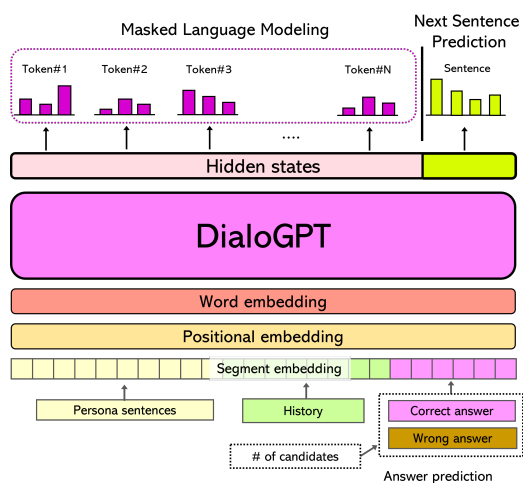


Figure 2: Overview of persona-based free talking model in FREETALKY.

Persona-based Free Talking Dialogue According to Grice’s maxims, there are several maxims that need to be followed to ensure a good conversation (Dale and Reiter, 1995; Looi and See, 2012). They are maxims of quantity, quality, relation, and manners.

To reflect these maxims, FREETALKY adopts a persona-based dialogue model. By assigning a persona to the system, it can provide consistent and friendly responses to the users. The users can practice conversing in English with the NAO robot similarly as they would converse with people. Additionally, they can set up the persona of the robot such that it includes their various interests or conversational topics.

Most existing foreign language learning systems provide only a single-turn dialogue. However, we deployed a multi-turn free-talking model using the pre-trained language model (Zhang et al., 2019) for a more realistic conversation. It was trained on a massive dialogue corpus to allow the model to utilize the learned dialogue patterns and construct colloquial expressions. We adopted PERSONA-CHAT (Zhang et al., 2018) as the training data for transfer learning of the model.

In the transfer-learning phase, the concatenations of persona sentences, dialogue history, and the user’s utterances were used as the input. With this input, the model learned the patterns inherent in dialogue data. As shown in Figure 2, the model was trained with two objective functions: next sentence prediction and masked language modeling. With the masked language modeling, the model learned next token generation in an auto-regressive manner. Simultaneously, it learned which answer is suitable for the next utterance among 20 candidates, using the next sentence prediction function.

The evaluation metrics for the model are hits@1 and perplexity. Hits@1 is used to measure the accuracy of the correct answer being chosen among 20 candidate answers. Perplexity is a metric for evaluating a language model for its ability to predict the next token. Low perplexity indicates low congestion and high performance of language models. In the results obtained from the qualitative analysis, hits@1 was 78.68 and perplexity was 14.19, when evaluated on the validation set of the PERSONA-CHAT dataset (Zhang et al., 2018). In addition, to handle ethical issues (e.g. hate speech, offensive words) which frequently arises from the language models due to the biases in toxic language datasets (Sap et al., 2019), we regenerate the sequence, if the generated sequence has certain words in our pre-built list of offensive words.

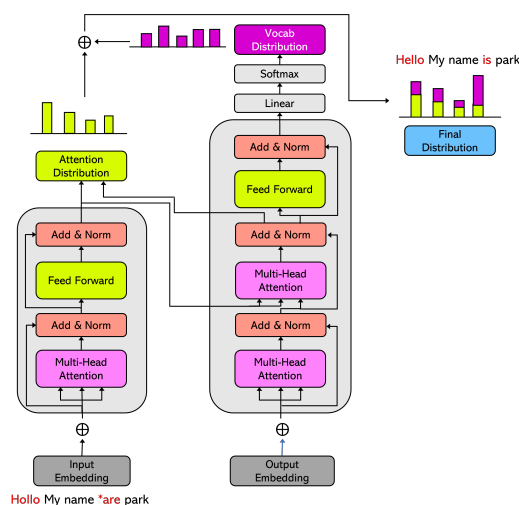


Figure 3: Overview of grammar error correction model in FREETALKY.

Grammar Error Feedback Several studies have proven the positive effects of feedback on error correction in foreign language learning (Lyster and Ranta, 1997; Loewen, 2013; Nassaji, 2016; Park et al., 2021). Inspired by these results, this study adapted the GEC model for FREETALKY to provide feedback on grammatical errors to users.

In terms of machine translation, GEC is a task in which the erroneous source sentence is translated to

the correct target sentence. That is, the sequence-to-sequence (S2S) model can be interpreted as noise sequence to denoise sequence (NS2DS (Park et al., 2020a; Park et al., 2020b)). All S2S models, such as CNN (Gehring et al., 2017), RNN (Bahdanau et al., 2014), and Transformer (Vaswani et al., 2017), can be utilized as NS2DS. In addition, existing rule-based and statistical models can be incorporated with NS2DS models in a hybrid form.

In this study, we adopted the vanilla Transformer (Vaswani et al., 2017) with a copy mechanism (Gu et al., 2016), which was suitable for the service owing to its lightweight structure. We applied the copy mechanism based on the characteristics of the training data in the GEC, where the source and target sentences had the same character set. Thus, we handled the out-of-vocabulary (OOV) problem using the copy mechanism in a principled manner.

The overall model structure is shown in Figure 3. It was deployed on a separate server and connected to FREETALKY by means of REST API using Flask. To enable scaling-up of the service, we additionally deployed the CPU-based model, as well as the GPU-based model (Klein et al., 2020). Through this function, explicit instructions are made available to the user.

4. FREETALKY for Xenoglossophobia

Xenoglossophobia, also known as foreign language anxiety, is a consistent and irrational fear revolving around foreign languages (Abinaya, 2016). People who suffer from xenoglossophobia experience anxiety while having conversations with people who speak a foreign language. There does not exist a specific method for treating xenoglossophobia in particular; however, the treatments suggested for general phobias can be utilized.

Most people who have a phobia are treated with exposure therapy. It is a method that involves exposing the patients to the situation, thus restricting the anxiety for a certain period. Though anxiety can temporarily increase during the treatment, it helps to release it in the long term, when the patients face real-world situations. Motivated by this treatment, we experimented to determine whether FREETALKY could practically help people who have xenoglossophobia. The NAO robot replaces the second human in the conversation, thus possibly relieving the anxiety of the user, and simultaneously makes the situation more realistic using the persona-based free talking system.

5. Qualitative Analysis

We analyzed the conversation between the user and FREETALKY in a qualitative manner. An example conversation is presented in Table 1. As shown in line (01), the FreeTalky system provides a randomly chosen persona through the selection of a number between 1 to 3. This setting enables virtual experiences in various situations where the users can experience

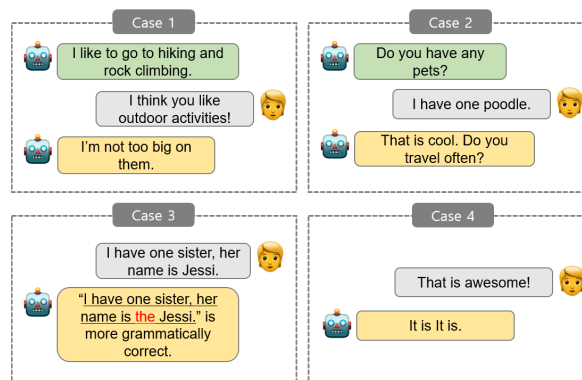


Figure 4: Four cases of errors. Case 1 is an erroneous sentence that conflicts with what the model said before. Case 2 is an unnatural utterance that suddenly changes the conversation topic. Case 3 is an error of overcorrection. Case 4 is a generic problem of generative models.

the feeling of talking to different people. In lines (08) and (14), FREETALKY generates persona-based answers. Its multi-turn conversation not only seems natural, but also continues until the user ends the conversation themselves. Moreover, it has the ability to ask an additional question from the user, as in (08). Once the conversation is over, it provides grammatical error feedback to the user, as in the last utterance of FREETALKY in Table 1.

Error Analysis Although the proposed system can hold a natural multi-turn dialogue and provide grammatical error feedback, several errors remain in need of improvement besides the given conversation. We divided the error types into four cases, as shown in Figure 4. The first case is the error of unnatural flow of conversation. Although the model attempts to converse about specific topics, it says things that conflict with what was said before, or with its persona, as shown in line (12) of Table 1. The second case is the utterance pattern learned from the pre-trained model. As we utilized the pre-trained model, the pattern learned from the large dialogue corpus tends to appear, as a result of which, the model sometimes changes the conversation topic and talks about itself. In Figure 4, the FreeTalky model suddenly changes the topic of conversation to travel. The third case is overcorrection. Overcorrection is a phenomenon wherein a grammatically correct input sentence is regarded as a subject that requires correction, and the sentence structure is disturbed. In Figure 4, the model changes a correct sentence of the user. The fourth case is a generic problem of generative models, that is, a model tends to generate tokens repeatedly and leads to a [unk] token issue, as shown in Figure 4.

6. FREETALKY Satisfaction Survey

We validated the efficiency of the proposed system through five middle school students, filtered by measurement (Al-Saraj, 2014), who experience foreign lan-

Example conversation with FREETALKY		
FreeTalky:	Hi, I am NAO. Please choose the number between 1 to 3 and tell me, then I will set my personality and tell you about my persona!	(01)
User:	Two	(02)
FreeTalky:	Persona setting is done. Let me introduce myself. I live in colorado. I like to go hiking in the spring. My favorite activity is rock climbing. I am a mechanical engineer.	(03)
FreeTalky:	Now you can have a conversation with me! If you want to finish the conversation, just say bye.	(04)
User:	Hi how are you?	(05)
FreeTalky:	I am doing great! I just got back from rock climbing. What are you up to?	(06)
User:	Wow... That's awesome. I am going to study English.	(07)
FreeTalky:	I am in colorado, we are going to go hiking sometime soon.	(08)
User:	I want visit Colorado. I think you like outdoor activities!	(09)
FreeTalky:	I do, but I'm not too big on them.	(10)
User:	I see. You.. engineer?	(11)
FreeTalky:	Yes, I am a mechanical engineer.	(12)
User:	Okay.. It's nice to meet you. Bye.	(13)
FreeTalky:	Conversation is done. Do you want to get grammatical feedback?	(14)
User:	Yes please.	(15)
FreeTalky:	Okay! You said, "I want visit Colorado", but "I want to visit Colorado", is a more grammatically correct expression. You said, "You engineer?", but "Are you an engineer?" is a more grammatically correct expression.	(16)
		(17)
		(18)
		(19)
		(20)

Table 1: An example conversation between the user and FreeTalky system. Users can set persona by choosing number and talk with FreeTalky system. It gives grammar error feedback to the user after chatting. Full video is attached in *Supp.*. The highlighted text indicates the user's grammar error and FreeTalky's feedback sentence.

guage anxiety. Further, we used six different questionnaires for system validation based on the Likert scale (Likert, 1932) that rates situations from 1 (*i.e.*, strongly disagree) to 5 (*i.e.*, strongly agree), as shown in Table 2.

As shown in Figure 5, most questions about the system were scored as strongly agree or agree. This result implies that the proposed system is suitable for alleviating people's xenoglossophobia.

Questionnaire	
Q1	I feel comfortable while interacting with a robot in a real human-like conversation.
Q2	I do not feel scared while talking to a robot.
Q3	Do you think FREETALKY will help alleviate your foreign language anxiety?
Q4	Are you satisfied with the performance of grammatical error feedback?
Q5	Do you think FREETALKY performs naturally in a multi-turn conversation?
Q6	Do you think persona is well applied in FREETALKY?

Table 2: Questionnaire of FREETALKY satisfaction survey.

Furthermore, we conducted Spearman correlation (de Winter et al., 2016) on the survey results in Figure 6. The results show a positive correlation in all question relationships. There is a significant ($p < 0.05$) positive correlation between Q1 and Q4, from Q2 to Q3, Q5, Q6, and between Q5 and Q6. Interestingly, Q3 and Q5, Q6 show a perfect correlation, which implies that there

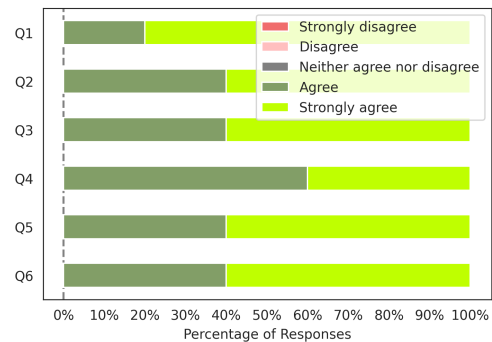


Figure 5: Response result of satisfaction survey of FREETALKY.

is a high association between the alleviation of foreign language anxiety and natural multi-turn dialogue and persona-applied conversation. These results prove that FREETALKY is appropriate for our objectives, that applying the persona and multi-turn dialogue to the conversation would result in the system being more user-friendly. Another highlight of the results corresponding to our objectives is that the humanoid robot NAO performed well in alleviating xenoglossophobia, and is user-friendly according to Q2, Q3, Q5, and Q6. Additionally, it demonstrated that GEC contributes to the comfort of the users while interacting with humanoid robots, according to the results of Q1 and Q4. Consequently, these factors determine the system satisfaction.

User interview before chatting with FREETALKY.	
Interviewer	Hi, nice to meet you.
User	I heard that you experience fear while meeting and conversing with foreigners.
Interviewer	That is right. I have learned English, but when I meet foreigners, I feel nervous, so I am unable to talk to them.
User	Okay, do you face any problems while reading and writing in English?
Interviewer	Yes, reading text in English is not a problem, but I get very nervous when I speak in English.
User	Have you made any efforts to overcome this?
Interviewer	I intended to go to an English conversation academy, but the cost was expensive, and I do not have enough time to go there.
User	Okay, then how about practicing conversing in English with a robot, instead of a human?
Interviewer	A robot? I have not talked to a robot before, so I do not know.
User	Then, why don't we try and see if it can help relieve your fear now?

User interview after chatting with FREETALKY.	
Interviewer	How was it, talking with the robot in English?
User	It was good because I felt more relaxed than when I talk to a real person. Also, it was good that this robot gave me feedback on grammatically wrong expressions. The robot had various voices, so I felt like it was a real situation.
Interviewer	You mean that talking with the robot helps relieve your psychological fear.
User	Yes, it does.
Interviewer	Great then, we will attempt to further improve this system to assist with your anxiety. Thank you
User	Thank you.

Table 3: User interview and experience of chatting with FREETALKY. As the user interview was conducted in Korean, the translated version is attached. The highlighted text indicates the notable points in the interview (see details in discussion).

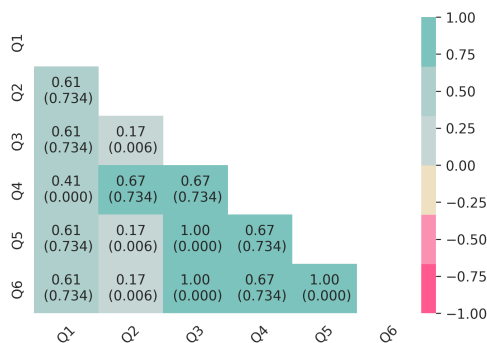


Figure 6: Correlation result of satisfaction survey of FREETALKY. Each cell indicates a correlation value, and numbers in parentheses are p-value (p-value<0.05 indicates statistical significance).

7. Discussion

Following the system satisfaction survey, we interviewed the actual users to determine an in-depth usability of the system besides the survey. The interview process aided the system development, as a predefined questionnaire could not handle unexpected answers, such as a narrative context. The interview transcript is presented in Table 3. The original interview video is also attached in the following link⁴.

⁴<https://youtu.be/P6Xyv3pi6F4>

8. Conclusion

We proposed FREETALKY, where users can receive feedback on English grammar errors and improve their English conversation skills. In addition, as FREETALKY is based on a persona dialogue system, the content of multi-turn conversations can be adaptively altered to suit the users and provide diverse conversational experience to the users. Furthermore, it is possible to alleviate the psychological pressure experienced by people suffering from xenoglossophobia, using the NAO robot.

9. Acknowledgements

This research was supported by Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2020-0-00368, A Neural-Symbolic Model for Knowledge Acquisition and Inference Techniques) and supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2021R1A6A1A03045425).

10. Bibliographical References

- Abinaya, R. (2016). Xenoglossophobia: Language learner’s perspective on learning. *American College Journal of English Language and Literature*, 5:1–4.
- Al-Saraj, T. M. (2014). Revisiting the foreign language classroom anxiety scale (flicas): The anxiety of female english language learners in saudi arabia. *L2 Journal*, 6(1).
- Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Dale, R. and Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive science*, 19(2):233–263.
- Damari, R. R., Rivers, W. P., Brecht, R. D., Gardner, P., Pulupa, C., and Robinson, J. (2017). The demand for multilingual human capital in the us labor market. *Foreign Language Annals*, 50(1):13–37.
- de Winter, J. C., Gosling, S. D., and Potter, J. (2016). Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, 21(3):273.
- Donmus, V. (2010). The use of social networks in educational computer-game based foreign language learning. *Procedia-Social and Behavioral Sciences*, 9:1497–1503.
- Gehring, J., Auli, M., Grangier, D., Yarats, D., and Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1243–1252. JMLR. org.
- Gu, J., Lu, Z., Li, H., and Li, V. O. (2016). Incorporating copying mechanism in sequence-to-sequence learning. *arXiv preprint arXiv:1603.06393*.
- Ha, J.-W., Nam, K., Kang, J., Lee, S.-W., Yang, S., Jung, H., Kim, E., Kim, H., Kim, S., Kim, H. A., et al. (2020). Clovacall: Korean goal-oriented dialog speech corpus for automatic speech recognition of contact centers. *arXiv preprint arXiv:2004.09367*.
- Horwitz, E. K., Horwitz, M. B., and Cope, J. (1986). Foreign language classroom anxiety. *The Modern language journal*, 70(2):125–132.
- Klein, G., Zhang, D., Chouteau, C., Crego, J. M., and Senellart, J. (2020). Efficient and high-quality neural machine translation with openmt. In *Proceedings of the Fourth Workshop on Neural Generation and Translation*, pages 211–217.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Loewen, S. (2013). The role of feedback. In *The Routledge handbook of second language acquisition*, pages 42–58. Routledge.
- Looi, Q. E. and See, S. L. (2012). Applying politeness maxims in social robotics polite dialogue. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 189–190.
- Lyster, R. and Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in second language acquisition*, 19(1):37–66.
- Nassaji, H. (2016). Anniversary article interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research*, 20(4):535–562.
- Park, C., Kim, K., Yang, Y., Kang, M., and Lim, H. (2020a). Neural spelling correction: translating incorrect sentences to correct sentences for multimedia. *Multimedia Tools and Applications*, pages 1–18.
- Park, C., Yang, Y., Lee, C., and Lim, H. (2020b). Comparison of the evaluation metrics for neural grammatical error correction with overcorrection. *IEEE Access*, 8:106264–106272.
- Park, C., Seo, J., Lee, S., Lee, C., Moon, H., Eo, S., and Lim, H.-S. (2021). Bts: Back transcription for speech-to-text post-processor using text-to-speech-to-text. In *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, pages 106–116.
- Sap, M., Card, D., Gabriel, S., Choi, Y., and Smith, N. A. (2019). The risk of racial bias in hate speech detection. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 1668–1678.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., and Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*.
- Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J., and Dolan, B. (2019). Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*.